don, K. Melhorn, T. Papatheodorou and P. Spirakis, ed. New York: Springer-Verlag, vol. 227, 1985, pp. 121-132.
[11] S. Muroga, *Logic Design and Switching Theory*. New York: Wiley, 1979.
[12] R. Nair *et al.*, "Linear time algorithm for optimal CMOS layout," in *VLSI: Algorithms and Architectures*. Amsterdam, The Netherlands: North-Holland, 1985, pp. 327-338.
[13] T. Uehara and W. M. VanCleemput, "Optimal layout of CMOS functional arrays," *IEEE Trans. Comput.*, vol. C-30, pp. 305-312, May 1981.
[14] S. Weimer, R. Y. Pinter, and J. A. Feldman, "Optimal chaining of CMOS transistors in a functional cell," in *Proc. Int. Conf. on Computer-Aided Design*, 1986, pp. 66-69.
[15] T. Lengauer and R. Muller, "Linear algorithms for optimizing the layout of dynamic CMOS cells," *IEEE Trans. Circuits Syst.*, vol. 35, pp.279-285, Mar. 1988.
[16] C. T. Mullen and R. H. J. M. Otten, "Layout compilation of linear transistor arrays," in *Proc. IEEE Int. Symp. on Circuits and Systems*, 1985, pp. 5-7.
[17] D. D. Gajski, *Silicon Compilation*. Reading, MA: Addison-Wesley, 1988.
[18] A. Hashimoto and J. Stevens, "Wire routing by optimizing channel assignment within large apertures," in *Proc. 8th ACM/IEEE Design Automation Conf.*, 1971, pp. 155-169.

Group Delay as an Estimate of Delay in Logic

J. Vlach, J. A. Barby, A. Vannelli, T. Talkhan, and C. J. Shi

***Abstract*—It is an accepted practice in signal delay estimation to model MOS digital circuits as *RC* circuits. In most cases Elmore's delay definition is applied.**

This paper has several objectives. First, it shows that Elmore's definition is *exactly* equivalent to the group delay of the network at zero frequency. Second, it presents a computationally efficient noniterative method to calculate this delay for networks with any linear elements and arbitrary topology. Third, it shows that in *RC* networks, under certain conditions, the Elmore delay and the 50% unit step response delay are related by a constant which is largely independent of the element values and topology. Finally, it presents an efficient method to obtain sensitivities of the delay with respect to any element in the network. These last two properties make design for specified delays feasible and computationally efficient.

I. Introduction

In recent years, a number of researchers addressed the problem of delays in logic and several authors, for instance [1]–[5], based their work on the paper by Elmore [6]. A recent study gives additional references in [7]. Considerable attention was also given to the simulation of devices by *RC* networks [8].

Originally Elmore's paper dealt with the first moment of the time response and all studies published so far were based on his arguments. Most of the methods were restricted to very simple topologies. A method based on the first moment but suitable for arbitrary topologies was published only recently [9]. Mathematical results in [9] turn out to be equivalent to our development. The difference is not so much in the mechanics of computations, but in the underlying understanding. Group delay is a well-known concept in filter design and many conclusions, elusive in the concept of time-domain moments, are easily reached from the point of view of group delay.

We show that Elmore's delay, T_D, is *exactly* equivalent to the group delay at zero frequency, $\tau(0)$, of a linear network having arbitrary topology and arbitrary linear elements. The equivalence of the two definitions is shown in Section II. An efficient numerical method for delay evaluation is developed in Section III for networks which can have arbitrary loops or interconnections. Such loops, as well as floating capacitors and grounded resistors, presented considerable problems in earlier developments. In [2], the authors had to resort to iterations in order to obtain the delay for networks with loops.

When time-domain solutions are feasible, then another widely accepted delay measure, D, uses a unit step at the input and calculates the time needed for the output to reach half of its final value. This definition is much better than Elmore's, but is also much more expensive, and it would be nice to establish a simple relationship between the two measures. This is indeed possible if the signal source and the output are at opposite ends of the network. In such case the 50% unit step delay, D, and the zero frequency group delay, $\tau(0)$, are related. This is discussed in Section IV.

Once we have found a reasonable relationship between the Elmore's definition and the 50% step response, we can use the former for design by optimization. For this we need sensitivities with respect to the elements of the network. A computationally efficient method is derived in Section V. Finally, Section VI outlines an optimization model in which the method can be used to design logic for specific values of delays.

II. Elmore's Definition and Group Delay

Elmore considers the network transfer function in the form

$$F(s) = K\frac{1 + a_1 s + a_2 s^2 + a_3 s^3 \cdots}{1 + b_1 s + b_2 s^2 + b_3 s^3 \cdots} \tag{1}$$

and defines the delay as

$$T_D = b_1 - a_1. \tag{2}$$

The group delay, a function of ω, is defined as

$$\tau(\omega) = -\frac{\partial \phi(\omega)}{\partial \omega} \tag{3}$$

with $\phi(\omega)$ being the phase characteristic of the transfer function. If we substitute $s = j\omega$ into (1), then

$$F(j\omega) = K\frac{(1 - a_2\omega^2 + \cdots) + j(a_1\omega - a_3\omega^3 + \cdots)}{(1 - b_2\omega^2 + \cdots) + j(b_1\omega - b_3\omega^3 + \cdots)}$$
$$= \frac{N_1 + jN_2}{D_1 + jD_2}$$

and the phase response is

$$\phi = \arctan\frac{N_2}{N_1} - \arctan\frac{D_2}{D_1}. \tag{4}$$

Consider only the denominator,

$$\frac{d}{d\omega}\left[\arctan\frac{D_2}{D_1}\right] = \frac{D_2' D_1 - D_2 D_1'}{D_1^2 + D_2^2}.$$

Manuscript received July 17, 1989. This paper was recommended by Associate Editor A. E. Ruehli.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

IEEE Log Number 9143296.

0278-0070/91/0700-0949$01.00 © 1991 IEEE

For $\omega = 0$ we have $D_2 = D_1' = 0$, $D_1 = 1$, $D_2' = b_1$, and the contribution of the denominator to the group delay at zero frequency is b_1. Applying the same steps to the numerator we find that the overall delay is identical to (2):

$$\tau(0) = b_1 - a_1 = T_D. \tag{5}$$

Group delay has been used extensively in the design of analog filters, and a considerable amount of knowledge is available about mutual relationship of various network responses. It is well known that peaks in the group delay response lead to ringing of the step response. If the group delay is flat or monotonically decreasing, the time-domain unit step response will not have an overshoot. In such case, there should be a reasonable correlation between the group delay at zero frequency and the actual delay measured at the half point of the final value.

As a test we designed a 25-th order LC filter with resistive terminations. It had maximally flat group delay in the passband and equi-ripple response in the stop band. Poles of such a filter are complex conjugate. The delay at 50% of the time-domain response is exactly equal to the group delay at zero frequency, and there is no overshoot in either the time or group delay responses. *RC* networks, used for simulation of the delay in MOS digital networks, will have all poles on the negative real axis. They will still have no overshoot in the time-domain response, but the group delay will be dropping at higher frequencies. The 50% unit step response delay, D, and T_D will not be equal, as in the above filter, but should be related by means of an approximately constant coefficient. This assumption, confirmed in the following, was possible only through the group delay argument. It was not available in the original concept of time-domain response moments.

III. Numerical Evaluation of Group Delay

Consider Fig. 1 in which one floating capacitor, C_5, and one grounded conductance, G_5, are added for generality. The input voltage is 1. The *KCL* equations in matrix form are

$$\boldsymbol{TX} = \boldsymbol{W} \tag{6}$$

where

$$\boldsymbol{T} = \boldsymbol{G} + s\boldsymbol{C} \tag{7}$$

where $\boldsymbol{X}$ is the vector of nodal voltages and $\boldsymbol{W}$ is the vector expressing the influence of the signal source. For $s = 0$, (6) simplifies to

$$\boldsymbol{GX}_0 = \boldsymbol{W} \tag{8}$$

the subscript denoting restriction to dc. For the example

$$\boldsymbol{G} = \begin{bmatrix} G_1 + G_2 + G_4 & -G_2 & 0 & -G_4 \\ -G_2 & G_2 + G_3 & -G_3 & 0 \\ 0 & -G_3 & G_3 & 0 \\ -G_4 & 0 & 0 & G_4 + G_5 \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} C_1 & 0 & 0 & 0 \\ 0 & C_2 + C_5 & -C_5 & 0 \\ 0 & -C_5 & C_3 + C_5 & 0 \\ 0 & 0 & 0 & C_4 \end{bmatrix}$$

and

$$\boldsymbol{W} = [G_1 \ \ 0 \ \ 0 \ \ 0]^T.$$

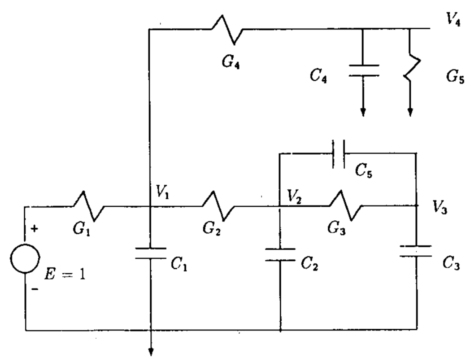


Fig. 1. A small *RC* network for explanation.

Define a vector, $\boldsymbol{d}$, which selects the desired output variable, and express the output as the product

$$F = \boldsymbol{d}^T\boldsymbol{X} \tag{9}$$

or for dc

$$F_0 = \boldsymbol{d}^T\boldsymbol{X}_0. \tag{10}$$

In the example, if output is taken at node 3, then

$$\boldsymbol{d} = [0 \ \ 0 \ \ 1 \ \ 0]^T.$$

For $s = j\omega$, the output is a complex number and can be written as

$$F(j\omega) = |F(j\omega)|e^{j\phi(\omega)}.$$

Taking the logarithm and differentiating with respect to ω:

$$\frac{1}{F}\frac{\partial F}{\partial \omega} = \frac{1}{|F|}\frac{\partial |F|}{\partial \omega} + j\frac{\partial \phi}{\partial \omega}.$$

Thus the group delay is

$$\tau(\omega) = -\frac{\partial \phi}{\partial \omega} = -\text{IM}\left[\frac{1}{F}\frac{\partial F}{\partial \omega}\right].$$

Since $s = j\omega$ and $(\partial s/\partial \omega) = j$, we can apply the chain rule and write

$$\frac{\partial F}{\partial \omega} = \frac{\partial F}{\partial s}\frac{\partial s}{\partial \omega} = j\frac{\partial F}{\partial s}.$$

This simplifies the final form for the group delay to

$$\tau(\omega) = -\text{RE}\left[\frac{1}{F}\frac{\partial F}{\partial s}\right] \tag{11}$$

and at dc

$$\tau(0) = -\frac{1}{F_0}\frac{\partial F_0}{\partial s}. \tag{12}$$

To obtain $\partial F/\partial s$, differentiate (6) with respect to s; this leads to

$$\boldsymbol{T}\frac{\partial \boldsymbol{X}}{\partial s} = -\frac{\partial \boldsymbol{T}}{\partial s}\boldsymbol{X} \tag{13}$$

which would normally be solved by LU decomposition. The result is substituted into the derivative of (9):

$$\frac{\partial F}{\partial s} = \boldsymbol{d}^T\frac{\partial \boldsymbol{X}}{\partial s} = -\boldsymbol{d}^T\boldsymbol{T}^{-1}\frac{\partial \boldsymbol{T}}{\partial s}\boldsymbol{X} = (\boldsymbol{X}^a)^T\frac{\partial \boldsymbol{T}}{\partial s}\boldsymbol{X}. \tag{14}$$

Here the vector $\boldsymbol{X}^a$ is obtained as the solution of

$$\boldsymbol{T}^T\boldsymbol{X}^a = -\boldsymbol{d}. \tag{15}$$

For dc, the expressions simplify to

$$\frac{\partial F_0}{\partial s} = (X_0^a)^T C X_0 \quad (16)$$

where X_0^a is obtained as the solution of

$$G^T X_0^a = -d. \quad (17)$$

In summary, one must solve (8) and (17) to get X_0 and X_0^a, insert into (10) and (16) and finally into (12).

Considerably simplifications follow if the network is reciprocal, because then $T^T = T$, $G^T = G$ and $C^T = C$. If all the capacitors are grounded

$$C = \text{diag}\,[C_i] \quad (18)$$

and (16) simplifies to

$$\frac{\partial F_0}{\partial s} = \sum_{i=1}^{N} C_i X_i X_i^a \quad (18)$$

where N is the size of the matrix. If, in addition, there are no resistors to ground, then all node voltages at zero frequency are equal to the source voltage, $X_i = 1$, $F_0 = 1$, and the group delay at dc reduces to

$$\tau(0) = \sum_{i=1}^{N} C_i X_i^a. \quad (19)$$

In this case, the evaluation requires only one solution of (17) and the use of (19). There are still no restrictions on the graph of the network and arbitrary loops can be present.

IV. Experiments

Usefulness of the Elmore's delay will be increased if there is a simple relationship between $T_D = \tau(0)$ and D. We first make a few observations. RC networks used for delay simulation transfer dc from the source to all nodes; the networks exhibits low-pass properties. Based on experience with passive analog filters we know that at internal points the responses have no resemblance to the overall response from input to output. It follows that attempts to apply the measure to internal nodes, especially to nodes close to the source, will be futile. For these reasons our study puts emphasis on input–output relationships and not on internal nodes.

Consider Fig. 2; in each experiment discard everything to the right of the node where the output is taken and calculate the ratio of the 50% step response delay and the group delay at zero frequency:

$$\beta = \frac{D}{\tau(0)}.$$

Two cases are summarized in Table I. As can be seen, β converges rapidly to a number approximately equal to 0.76, although the time constants are vastly different. This led us to the idea that a similar situation will exist in more complicated networks. We tested many complicated RC networks and grounded resistors as well as floating capacitors were allowed. Each network configuration was run with many different element values. To establish some regularity, we set, for instance, all capacitors with even subscripts to one value and capacitors with odd subscripts to another value such that the ratios C_{even}/C_{odd} were 1, 10 or 100, etc. Overall, more than 80 cases were tested and tables of some of the results are in [10]. The experiments took some internal nodes that were far from the source. Our largest network had 80 elements, many loops, and a floating capacitor and a grounded resistor at every node.

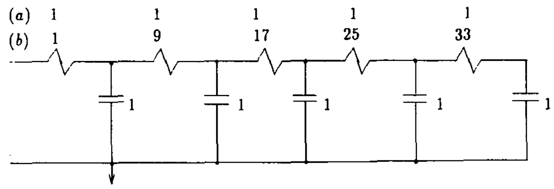


Fig. 2. RC ladders with different time constants. For calculation at any node the rest of the network to the right is discarded.

TABLE I
Comparison of β for the Networks in Fig. 2

	1	2	3	4	5
β for resistors 1, 1, 1, 1, 1	0.693	0.742	0.756	0.756	0.756
β for resistors 1, 9, 17, 25, 33	0.693	0.723	0.758	0.767	0.771

Under the conditions specified above, we found that D and $\tau(0)$ are related in RC networks. Using 50 cases with outputs at the most distant nodes from the source we obtained the average ratio

$$\beta = \frac{D}{\tau(0)} = 0.7533$$

with a standard deviation of β equal to 0.0229.

In Fig. 3 we give one example with many loops. The capacitors were changed as indicated and the resulting delays and their ratios are also given. Another example, Fig. 4, considers two parallel RC lines running close to each other, a test useful for VLSI interconnect delay. All $R_i = 1$, all grounded $C_i = 1$, and all coupling capacitors, C_f, were equal. In Fig. 4(a), all nodes of the second line were grounded and thus C_f were connected to ground. In Fig. 4(b), the second line was connected as shown. The results are collected in Table II for several ratios of C_f/C.

Summarizing the experimental results, we conclude that the Elmore delay can be used fairly reliably if the input and output are far from each other. Under these conditions one can obtain a good estimate of the 50% time delay by using the expression $D = 0.7533\tau(0)$.

V. Sensitivity of Group Delay to Network Elements

The group delay theory lends itself to an easy evaluation of sensitivities with respect to network elements and thus to optimization by means of routines which rely on the gradients. To derive the necessary formulas we need the derivative of (12) with respect to any element of the network, h. Starting with (11):

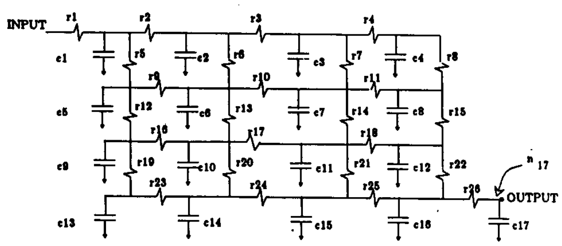
$$\frac{\partial \tau(\omega)}{\partial h} = \text{RE}\left[\frac{1}{F^2}\frac{\partial F}{\partial h}\frac{\partial F}{\partial s} - \frac{1}{F}\frac{\partial^2 F}{\partial h \partial s}\right]. \quad (20)$$

For dc we drop the RE symbol and use zero subscript. The derivative $\partial F/\partial h$ is obtained from (14) by writing h instead of s:

$$\frac{\partial F}{\partial h} = (X^a)^T \frac{\partial T}{\partial h} X. \quad (21)$$

For dc, this simplifies to

$$\frac{\partial F_0}{\partial h} = (X_0^a)^T \frac{\partial G}{\partial h} X_0. \quad (22)$$



C_{odd}	C_{even}	Group Delay [msec]	Exact Delay [msec]	$\frac{T_{exact}}{T_D}$
1.0×10^{-6}	1.0×10^{-6}	4.4462×10^{1}	3.27×10^{1}	0.735
1.0×10^{-6}	1.0×10^{-3}	2.086×10^{4}	1.55×10^{4}	0.743
1.0×10^{-3}	1.0×10^{-6}	2.404×10^{4}	1.8×10^{4}	0.748
1.0×10^{-4}	1.0×10^{-3}	2.3×10^{4}	1.7×10^{4}	0.739
1.0×10^{-3}	1.0×10^{-4}	2.589×10^{4}	1.95×10^{4}	0.753

Fig. 3. *RC* network with many loops.

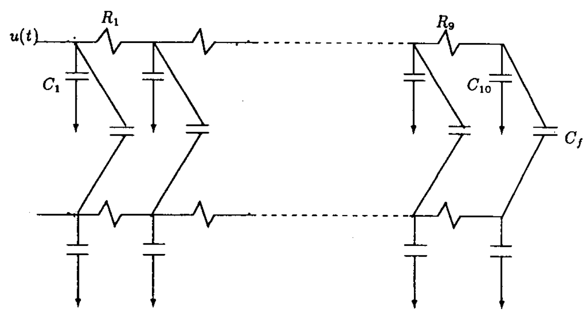


Fig. 4. Simulation of two parallel *RC* lines with capacitive coupling.

TABLE II
SIMULATION RESULTS FOR THE NETWORKS IN FIG. 4

$\frac{C_f}{C}$	$\tau(0)$	D Case (a)	D Case (b)	β Case (a)	β Case (b)
0	55.0	41.7	41.7	0.758	0.758
0.1	60.5	45.8	45.6	0.757	0.754
0.5	82.5	62.4	58.6	0.756	0.710

In (20) we still need the second derivative. It is obtained by differentiating (21) with respect to s:

$$\frac{\partial^2 F}{\partial h \partial s} = \frac{\partial (X^a)^T}{\partial s}\frac{\partial T}{\partial h}X + (X^a)^T\frac{\partial^2 T}{\partial h \partial s}X + (X^a)^T\frac{\partial T}{\partial h}\frac{\partial X}{\partial s}. \quad (23)$$

This simplifies considerably for dc:

$$\frac{\partial^2 F_0}{\partial h \partial s} = \frac{\partial (X_0^a)^T}{\partial s}\frac{\partial G}{\partial h}X_0 + (X_0^a)^T\frac{\partial C}{\partial h}X_0 + (X_0^a)^T\frac{\partial G}{\partial h}\frac{\partial X_0}{\partial s} \quad (24)$$

and the formula, in fact, splits into two formulas. If the element is a capacitor, then

$$\left.\frac{\partial^2 F_0}{\partial h \partial s}\right|_{h=C} = (X_0^a)^T\frac{\partial C}{\partial h}X_0. \quad (25)$$

If it is a conductance

$$\left.\frac{\partial^2 F_0}{\partial h \partial s}\right|_{h=G} = \frac{\partial (X_0^a)^T}{\partial s}\frac{\partial G}{\partial h}X_0 + (X_0^a)^T\frac{\partial G}{\partial h}\frac{\partial X_0}{\partial s}. \quad (26)$$

For these expressions, $\partial X/\partial s$ would normally be given by the solution of (13), but at dc it is obtained as the solution of

$$G\frac{\partial X_0}{\partial s} = -CX_0 \quad (27)$$

The derivative $\partial X^a/\partial s$ is obtained by first differentiating (15) with respect to s:

$$T^T\frac{\partial X^a}{\partial s} = -\frac{\partial T^T}{\partial s}X^a. \quad (28)$$

Simplification at dc reduces (28) to

$$G^T\frac{\partial X_0^a}{\partial s} = -C^TX_0^a. \quad (29)$$

The right-hand sides in (27) and (29) are easily generated, because the vectors X_0 and X_0^a are already available. In addition, transposition of C and G is not needed because the matrices will be symmetric. The G matrix is already LU decomposed from previous solutions and only two new forward-back substitutions are needed. Expressions (25) and (26) cost next to nothing computationally because a grounded element appears in the matrix in only one position and the whole vector-matrix-vector product reduces to just one multiplication. If the element is floating, its value appears in the matrix in four places and each vector-matrix-vector product reduces to two subtractions and one multiplication. If more than one output is needed, (29) must be resolved for each new output, but this will require only a new forward-back substitution. Equation (27) is solved only once.

VI. Design Application

Since derivatives of the group delay with respect to network elements are so easy to obtain, the method can be used for design by modifying values of some *RC* network elements. The resulting values can then be traced back to transistor sizes or other representations.

As a simple example, consider the case where a signal should propagate through an arbitrary structure, with the desired delays at m selected points to be d_1 through d_m. We can define the function

$$L = \frac{1}{2}\sum_{i=1}^{m}(\beta\tau_i(h_1, h_2, \cdots h_m) - d_i)^2$$

and apply minimization to the problem

$$\text{minimize } L$$
$$\text{subject } l_j < h_j < u_j. \quad (25)$$

Here l_j and u_j are lower and upper bounds on element values, respectively. The minimum, $L = 0$, will be reached when the delays are equal to the desired values, d_j. Optimization will need the gradient, which in this case is

$$\frac{\partial L}{\partial h_j} = \sum_{i=1}^{m}[\beta\tau_i(h_1, h_2, \cdots h_m) - d_i]\beta\frac{\partial \tau_i}{\partial h_j}. \quad (26)$$

The values of τ_i and $\partial\tau_i/\partial h_j$ are obtained from (12) and (20).

VII. Conclusion

Elmore's delay definition and group delay at zero frequency are two equivalent definitions but the group delay concept gives much more insight. It is also easy to calculate sensitivities of the group delay with respect to element changes and thus the method can be used for design. The paper establishes experimentally that in *RC* structures the delay measured at 50% of unit step input and group delay at zero frequency are related by a proportionality constant, which is practically independent of topology and element values if the source of signal and points of interest are at opposite ends of the network.

References

[1] J. Rubinstein, P. Penfield, and M. A. Horowitz, "Signal delay in *RC* tree networks," *IEEE Trans. Computer Aided-Design*, vol. CAD-2, pp. 202–211, July 1983.

[2] T. Lin and C. A. Mead, "Signal delay in general *RC* networks," *IEEE Trans. Computer-Aided Design*, vol. CAD-3, pp. 331–349, Oct. 1984.

[3] J. L. Wyatt, Jr., Q. Yu, C. A. Zukowski, H. Tan, and P. R. O'Brien, "Improved bounds on signal delay in MOS interconnect," in *Proc. 1985 IEEE Int. Symp. Circuits and Systems*, 1985, pp. 903–906.

[4] P. R. O'Brien and J. L. Wyatt, "Signal delay in ECL interconnect," in *Proc. IEEE Int. Symp. on Circuits and Systems*, 1986, pp. 755–758.

[5] P. K. Chan, "An extension of Elmore's delay and its application for timing analysis of MOS pass transistor networks," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 1149–1152, Nov. 1986.

[6] W. C. Elmore, "The transient response of damped linear networks with particular regard to wide-band amplifiers," *J. Appl. Phys.*, vol. 19, pp. 55–63, Jan. 1948.

[7] P. K. Chan and M. D. F. Schlag, "Bounds on signal delay in *RC* mesh networks," *IEEE Trans. Computer-Aided Design*, vol. 8, pp. 581–589, June 1989.

[8] J. L. Wyatt, "Signal propagation delay in RC models for interconnect," in *Circuit Analysis, Simulation and Design, Part 2*, A. E. Ruehli, Ed. Amsterdam, The Netherlands: North-Holland, 1987, pp. 254–291.

[9] L. T. Pillage and R. A. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Computer-Aided Design*, vol. 9, pp. 352–366, Apr. 1990.

[10] I. H. Talkhan, "Signal delay estimation in logic networks," M.Sc. thesis, Univ. Waterloo, Dep. Elec. Eng., Waterloo, Ont., Canada, 1986.